

CRISPR-UMI Step by Step: A protocol for robust CRISPR screening

Ulrich Elling (✉ Ulrich.elling@imba.ac.at)

Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr Gasse 3, 1030 Vienna, Austria.

Georg Michlits

Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr Gasse 3, 1030 Vienna, Austria.

Thomas R Burkard

Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Campus Vienna Biocenter 1, 1030 Vienna, Austria

Maria Novatchkova

Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Campus Vienna Biocenter 1, 1030 Vienna, Austria

Method Article

Keywords: CRISPR, Screen, Etoposide, iPS, reprogramming, UMI

Posted Date: October 17th, 2017

DOI: <https://doi.org/10.1038/protex.2017.111>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

CRISPR-UMI extends the existing repertoire of CRISPR-screening methods. It circumvents cell heterogeneity, a consequence of Cas9 genome editing, by scoring single cell derived clones individually. The strength of this new CRISPR screening method is its robustness towards clonal heterogeneity and clonal outliers and is therefore expected to be most useful in challenging biological screens with strong bottlenecks and clonal effects such as organoid or in-vivo screens. This step-by-step protocol is an addition to the publication CRISPR-UMI: Single cell lineage tracing of pooled CRISPR/Cas9 screens doi: 10.1038/nmeth.4466. It contains a detailed description for pooled CRISPR screening using CRISPR-UMI. It especially highlights the steps which are critical and unique to the use of CRISPR-UMI. Those critical steps are library preparation at very high complexity of up to 100million unique plasmids, and data analysis where unique guide-UMI pairs are evaluated separately.

Introduction

Pooled CRISPR screens are a powerful tool to assess gene function. However, conventional analysis suffers from cellular heterogeneity that is either a consequence of Cas9 editing or cell culture intrinsic. Here we present CRISPR-UMI (Unique Molecular Identifier), a single cell tracing approach, providing a robust screening method that can detect, and thus overcome, cellular heterogeneity and clonal outliers. For details see attached File: Main document 1 CRISPR-UMI Step-by-Step. To be found in the tab "Figures"

Reagents

- XbaI (NEB) - MfeI (NEB) - EcoRI (NEB) - rSAP (NEB) - Gel Extraction Kit (QIAGEN) - PCR Purification Kit (QIAGEN) - Phusion® Polymerase (NEB) - T4 Ligase (NEB M0202) - XL-1 Blue Electrocompetent cells (Agilent) - SOB 20g/L Bacto Tryptone BD, 5g/L Yeast Extract BD, 10mM NaCl Merck, 2.5mM KCl Sigma Aldrich - NucleoBond® Xtra Maxi (Macherey-Nagel) - BbsI (NEB) - Phenol:Chloroform:Isoamylalcohol = 25:24:1 (Carl Roth) - Chloroform (Sigma Aldrich) - Platinum-E cells (Cell Biolabs RV-101) - RV-helper plasmid: pCMV-Eco Envelope Vector (Cell Biolabs RV-112) - HBS Buffer (280mM NaCl, 50mM HEPES, 1.5mM Na₂HPO₄, adjust pH to 7.00 with 0.5M NaOH) - Polybrene (Sigma-Aldrich) - G418 (Gibco) - Cell line: dox inducible Cas9 mouse embryonic stem cells AN3-12.9 - Embryonic stem cell medium (450 ml DMEM (Sigma D1152); 75 ml FCS (Invitrogen); 5.5 ml P/S (Sigma P0781); 5.5 ml NEAA (Sigma M7145); 5.5 ml LGlu (Sigma G7513); 5.5 ml NaPyr (Sigma S8636); 0.55 ml βME (Merck 805740; dilute 10ul βME in 2.85 ml PBS for a 1000x stock), 7.5ul LIF (Sigma; 2mg/ml)). - Doxycycline (Sigma-Aldrich) - 2X SDS Lysis Buffer: (10mM Tris-HCl pH 8.0, 5mM EDTA, 100mM NaCl, 2% SDS, add fresh directly before use: Proteinase K from a 10X stock 10mg/ml Proteinase K stored in 50% Glycerol at -20°C. To a final concentration of 1mg/ml) - RNaseA (100mg/ml) (Qiagen) - Phenol:Chloroform:Isoamylalcohol = 25:24:1 (Roth) - Chloroform (Sigma) - 5M NaCl solution (Sigma) - RNase A (QIAGEN) - TE, Tris-EDTA solution: (10mM Tris-HCl pH 8.0, 5mM EDTA) - SpeedBeads™ magnetic carboxylate modified particles (GE45152105050250 Sigma Aldrich) - PHUSION® polymerase (NEB) - KlenTaq polymerase (DNA Polymerase Technology) - Binding Buffer (

(20% PEG8000 \ (FLUKA) 2.5M NaCl) - Eppi-Magnet \ (for large quantities 15mL Falcon Magnet) - PCR Purification Kit \ (QIAGEN)

Equipment

Electroporation: Gene Pulser II \ (Biorad) NGS Sequencing: HiSeq2500 \ (Illumina) Tissue cell culture standard lab equipment Molecular biology standard lab equipment

Procedure

Chapter1: Library cloning Step 1: Vector design The Vector backbone contains ampicillin resistance for amplification in bacteria, viral packaging sequence Psi and long terminal repeats \ (LTRs) for generation of retrovirus. However, a lentiviral backbone can be equally used. The sgRNA cassette contains a U6 promotor and cloning site for CRISPR-guides \ (Step 4), improved CRISPR-Scaffold as described 2 and PGK NeoR for selection. Cloning of P5 and P7 Illumina adaptor sequences into the vector backbone allows direct Sequencing of the viral cassette. The essential modification for CRISPR-UMI is the integration of random sequences termed barcodes \ (barcodes in combination with sgRNA make the UMIs Unique molecular identifiers) and the illumina i7 \ ('index') primer binding site for barcode-sequencing. A PCR product reaching from illumina P5 Adaptor to illumina P7 adaptor can be used directly for next generation sequencing on an Illumina HiSeq2500 sequencer using dual indexing. Illumina's "read 1" is read with a custom primer \ (see Table 1 Primer_oligos.xls) and gives the CRISPR-guide sequence, "index1-read" gives the barcode sequence, and "index2-read" the experimental index to differentiate between samples \ (e.g. treated vs control, or replicas). A further modification is flanking the PCR-Amplicon with Pac-I restriction sites which enable enrichment of the integrated cassette from genomic DNA by performing size selective precipitation on magnetic beads. Library cloning is a two-step cloning process. First, random nucleotides \ (10nt) at a complexity of about 1 million later referred to as barcodes \ (BCs) are cloned in to the vector backbone \ (here referred to as step 2), then CRISPR-guides are cloned into that barcode-library \ (here referred to as step 3) reaching library complexities of 100 million due to combining barcodes and guides. Comment: The illumina i7 binding site is usually used for reading out the experimental index to differentiate between samples, we make use of illumina's dual-indexing approach were a second \ (in the case of CRISPR-UMI the only) experimental index can be read adjacent to the P5 adaptor. For details see attached Main Document 1 Step 2: Barcode-library cloning This library cloning step introduces random nucleotides of a length of 10bp downstream of Illuminas i7 primer binding site together with the P7 Adaptor into the vector backbone. While the theoretical complexity of a 10bp random sequence is limited to about 10^6 variations. Cloning complexity should be at least 1 million but higher complexities are desirable. For details see attached Main Document 1 Step 3: Guide selection sgRNAs targeting mouse nuclear genes as well as drugged orthologues and a set of hand selected genes with 4 sgRNAs per gene \ (5 sgRNAs per gene for the subset drugged genes) were selected by a bioinformatics pipeline. We aimed to design a guide selection algorithm taking both guide efficiency as well as biological effect due to gene structure into account. The basis of the guide selection is the activity

score as described by Doench et al.³. Additionally, we identified properties of each guide and exon under consideration and penalized the Doench score accordingly. We identified all exonic PAM sites in the mouse genome mm10⁴. We excluded sgRNAs that are incompatible with our cloning strategy (contain: GAAGAC, GTCTCC, CTCGAG, CGTCTC or GAGACG, start with: AAGAC or end with: CTCGA). We then calculated Doench-scores for all potential sgRNAs. We penalized the Doench-scores based on heuristic rules that aim to select sgRNAs which most likely lead to LOF phenotypes. Those rules include exon properties such as presence or absence of protein domains annotated in Pfam database⁵, exon size, and whether or not exon length is a multiple of 3bp. Then we created penalties for exon distribution, to spread sgRNAs over many exons where only the sgRNA with the best Doench score per exon does not get penalized. We also avoided sgRNAs that are less than 4nt away from another better scoring sgRNA. Furthermore, we penalized sgRNAs that cut DNA upstream of a possible alternative ATG start codon and sgRNAs that cut in exons that are not common to all annotated transcripts from that locus. We avoided sgRNAs that contain a stretch of 4 or more T in a row which would act as a Pol-III Terminator. We calculated a distance-penalty based on the distance from the sgRNA to the transcriptional start ranging from 1 to 0.5. Then we calculated a simple off-target prediction (see associated publication) against all exonic sequences containing a PAM site. The off-target prediction scores weight mismatches by position in the sgRNA sequence^{6,7}. We re-ranked the penalized Doench score including the off-target analysis and picked the top 4 sgRNAs per gene (the top 5 sgRNAs for Druggable genes) for chip oligo synthesis (CustomArray Inc.). For negative control guides we used a published list of human control guides⁸ and removed all guides which had a perfect match against the mouse genome. We included a total of 112 control guides into our mouse library targeting 6560 genes. For details see attached Main Document 1 Step 4: sgRNA cloning For CRISPR-UMI library cloning a complex insert (e.g. 26500 sgRNAs PCR amplified from chip-oligo synthesis) is cloned into a complex vector backbone (containing up to 1 million different UMIs). In every possible combination, this would allow a theoretical complexity of 26.5 billion (when using 26500 guides) unique CRISPR-UMI pairs. Cloning efficiency should be at least 1000X per guide (i.e. 30million for 30.000 guides) to generate a library complex enough for CRISPR-UMI. We aimed to generate Libraries of a complexity of about 85 million for 26500 guides. For details see attached Main Document 1 Chapter 2: Screening Step 1: Generation of virus in Plat-E cells We use Plat-E cells for packaging virus. Since the CRISPR-UMI plasmid library is of very high complexity (e.g. 85 million) and we want to keep complexity and even representation of individual guide-barcode pairs. We recommend to infect at least six 150mm dishes (about 250million cells) for a 26500 sgRNA library to retain the necessary complexity. For details see attached Main Document 1 Step2: Execution of the screen A basic principle to keep in mind are cell numbers that need to be (or can be) carried through the experiment. For example, if running the screen with 30.000 guides per gene we aim to always keep at least 30million cells in the experiment (1000x representation). We grow cells from 30million to 300million and split every 2nd day at a ratio of 1:10 (keep at least 30 million cells and discard the rest). CRISPR-UMI offers 2 variations: with or without limiting dilution - clonal expansion. With limiting dilution - clonal expansion: In this protocol CRISPR-UMI introduces an artificial bottleneck after CRISPR-gene editing has occurred. Depending on the screen setting introducing a strong bottleneck means discarding 95-99 % of all cells and then expanding the remaining 1-5% of cells. By doing so we reach cell numbers

much lower than the complexity of the CRISPR-UMI library and most cells in the experiment will carry a unique guide-barcode pair (UMI). Therefore, after expansion every UMI will carry a clonally selected uniquely repaired CRISPR cutting site. This contrasts with conventional CRISPR screens where cells carrying the same guide are heterogeneous in the way the CRISPR-cut was repaired. We recommend a limiting dilution and expansion for negative selection screens when comparing two conditions, because you can make use of multiple isogenic clones that you can compare in two settings. The cost of a limiting dilution is that the extra time required for expansion can cause shifts in representation and that under-represented guides can be lost completely from the experiment. Note that some experiments (like in-vivo screens with bottlenecks such as engraftment of cells or differentiation screens with moderate efficiency) introduce this “limiting dilution” step inherently. Comment: Why a limiting dilution generates isogenic clones: Assume a single cell is infected with a single virus carrying a unique UMI. Before and during gene-editing this will give rise to a handful of daughter cells which all carry the same guide but generate different mutations due to random mistakes in error-prone repair mechanisms. As a consequence daughter cells are heterogeneous like in a conventional CRISPR screen. By introducing a strong enough dilution step after CRISPR mutations are set, only one daughter cell will remain in the experiment, after expanding the population again the UMI will now be unique to all “grand-daughters” and in contrast to a conventional CRISPR screen, all grand-daughters will carry the same CRISPR-mutation. Note that during the dilution-expansion step most UMIs will be completely lost with not a single daughter cell remaining in the experiment. Without limiting dilution – clonal expansion: Not introducing a strong limiting dilution but still using CRISPR-UMI is also an option. While the benefit of isogenic clones is lost, you can still use UMIs as conceptual replicates and detect and exclude artefacts or outliers from data analysis. Also in positive selection where selection events are considered rare occasions, you can use UMIs to differentiate between incidence of an event (were the number of independent UMIs indicates the frequency of an event) and abundance (indicated by the counts per UMI which give information about the extent of the positive selection event). For details see attached Main Document 1 Step 3: Genomic DNA isolation, PCR amplification and next generation sequencing If cell numbers are not limiting, we recommend to harvest 3 fold more cells than the number of reads to be retrieved from NGS Sequencing. For 1 lane on a HiSeq2500 that gives about 250million reads we recommend to harvest 750million cells. More cells may be harvested as backups or frozen as live stocks. All quantities given in the protocol are for processing 750million cells. Realistically those 750million cells will be subdivided into different experimental conditions, but for the purpose of this protocol total quantities for processing 750million cells are given. For details see attached Main Document 1 Chapter 3: Data analysis Step 1: Assignment and counting of sequencing reads We use samtools, fastx-toolkit and bowtie to assign guides and experiments to sequencing reads and then count sequencing reads of UMIs. This section describes how we convert the bamfile from Illumina sequencing to a tab separated text file with the columns: Guidename Samplename (e.g. ctrl_1, treated_1) Barcode Sequence Read count In the later sections of this protocol this tab separated text file will be the input and starting point of more specialized analysis scripts. For details see attached Main Document 1 Step2: Negative selection, CRISPR-UMI pipeline The main purpose of this section of data analysis is to document and describe the scripts and calculations that were used to evaluate hits in a negative depletion setting. The key script in the analysis pipeline is

CRISPR_UMI.py. It prepares input files for MAGeCK 10 for both conventional CRISPR analysis (ignoring BCs) and CRISPR-UMI analysis in parallel. By running an algorithm called POPTOP(x) prior to analysis it also allows to remove a certain number of clones (x) per guide, always removing the clones with the highest read support (ctrl and treated condition taken together), prior to analysis. This analysis allowed us to show that some of the clones with highest read support are responsible for false positive signals in conventional CRISPR-screening and that CRISPR-UMI screening is robust towards those outliers. For CRISPR-UMI individual clones are evaluated using MAGeCK to give a depletion score for guides and CRISPR_UMI.py gives median depletion (reads treated/reads ctrl) for every guide. Combining those two values for every guide allows to robustly score the effect of each guide. The analysis starts from a tab separated text file with the columns: Guidename Samplename (e.g. ctrl_1, treated_1) Barcode Sequence Read count The analysis compares 2 experiments against each other using A) a conventional approach ignoring the clonal information provided by barcodes or B) using CRISPR-UMI analysis. For the conventional approach A) the sequence read counts for the same guide and the same sample but different BC Sequences are all added together and a file with 4 columns (Guidename, Genename, ctrl (reads), exp (reads)) is generated. This file serves as an input file for MAGeCK 10. For CRISPR-UMI analysis. B) the script calculates depletion by $RPM(ctrl)/RPM(treated)$ and determines the median depletion of clones for each guide. It generates a file with the median depletion for all guides. It also generates a file for MAGeCK analysis but with the 4 columns (UMI-name, Guidename, ctrl (reads), exp (reads)). Run MAGeCK on that file and the result will be a list of all guides ranked by MAGeCK's robust ranking algorithm. Combine the median depletion of every guide with the MAGeCK neg score (the score by which MAGeCK typically ranks genes). To rank genes we rank guides by median depletion and calculate $rank/(total\ number\ of\ guides)$, we then rank all guides by MAGeCK neg score and calculate $rank/(total\ number\ of\ guides)$. Multiplying those two values gives a score for every guide. We combine scores for every guide using Fisher's method to generate a depletion score for each gene. For details see attached Main Document 1 Step 3: Positive Selection; Incidence vs abundance analysis For positive selection screens CRISPR-UMI can be used to differentiate between Abundance (that is total read number of a guide) and Incidence (number of independent barcodes sequenced) as we demonstrated for a screen for roadblocks of reprogramming. For details see attached Main Document 1 Step 4: Clonal size estimation in reprogramming screen This Section describes scripts used for estimating colony size in a positive selection screen of reprogramming. Mouse embryonic fibroblasts were infected with CRISPR-UMI library and reprogrammed to induced pluripotent stem cells. In the example data set samples are labelled C1-C4 for controls (biological replica, MEFs from 4 mice) and E1-E4 for experiment (reprogramming of MEFs from those 4 mice), E1A and E1B, are technical replicas. In this section of the Step-by-step protocol the term UMI is used to solely describe the 10nt barcode and not the combination of a guide-barcode, this is not coherent with the rest of the Step-to-Step protocol or the associated publication. The colony size is reflected in reads per UMI and colony number in the number of different UMI per guide. The Analysis described here is an estimation of average colony size depending on the gene knocked out with CRISPR-UMI. For details see attached Main Document 1 Step 5: Comparisons of CRISPR-UMI vs conventional CRISPR screening We use two approaches to evaluate and quantify screen-quality. Both quality checks are carried out on guide level. One is signal to noise ratio, where we plot all

guides in a volcano plot and define signal as distance from the origin and noise as the standard deviation among non-targeting ctrl guides. The other method ranks all guides and calculates for how many guides per gene are found among the top 5,10,20,30, and so on guides. For details see attached Main Document 1

Timing

Library cloning: 1 week Screen procedure: 3 weeks DNA preparation and Sequencing: 1-2 weeks Data Analysis: 1 week

Anticipated Results

CRISPR-UMI offers improved signal to noise ratios in negative selection over conventional CRISPR screens and allows removal of single-cell derived outlier clones that generated putative false positive hits in a conventional CRISPR screen. (See associated publication).

References

1. Michlits, G. et al. CRISPR-UMI: Single cell lineage tracing of pooled CRISPR/Cas9 screens. *Nature Methods* doi:10.1038/nmeth.4466 (2017).
2. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491 (2013).
3. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* 32, 1262–1267 (2014).
4. Rosenbloom, K. R. et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 43, D670–81 (2015).
5. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–85 (2016).
6. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology* 32, 677–683 (2014).
7. SQ, T. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* 33, 187–197 (2014).
8. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* 343, 80–84 (2014).
9. Elling, U. et al. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* 9, 563–574 (2011).
10. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554 (2014).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement0.docx](#)
- [supplement0.zip](#)
- [supplement0.xlsx](#)

- [supplement0.zip](#)
- [supplement0.zip](#)